

UCS505
CONVERSATIONAL AI: DATA SCIENCE
Minor Project



**Document and QR Code Scanner
with Text Detection**

Submitted By:

Sargun Bindra (101903175)

Submitted To:

Dr Sahil Sharma

**Computer Science and Engineering Department
Thapar Institute of Engineering and Technology**

**Patiala-147004
May 2022**

Abstract

The scanner at your fingertips. Despite the enormous growth rate of digital information - the amount of digital information grew by 62 percent to 800 billion gigabytes in 2009 according to EMC- paper still offers unique advantages and will remain in daily use. As the paperless office has not yet arrived, the challenge of digitally capturing content of paper documents remains to be solved. Especially when trying to quickly capture pieces of text, diagrams or tables, traditional office devices such as the widespread personal desktop scanners or shared network scanners disrupt the user's workflow and are considered cumbersome and time consuming. In a pilot interview study with CIOs and IT managers of 9 large and international companies we identified a distinct need for an efficient personal solution for such small capture tasks. The solution to this problem is a mobile scanner which is available anytime and anywhere.

QR codes are two-dimensional matrix barcodes. Last decade witnessed a steady growth in commercial and business-oriented usage of these QR codes with the advent of smart and web capable mobile phones. The decoding procedure consists of image binarization, QR code extraction, perspective transformation and resampling, and error correction. By these steps, we can recognize different types of QR code images. QR code can be scanned from any direction because the proportion of position detection patterns is not changed with the scanning direction. QR code supports different encoding types and versions. We can choose an appropriate encoding type and version to reduce the size of QR code image. QR codes are frequently used to track information about products in a supply chain and – because many smartphones have built-in QR readers – they are often used in marketing and advertising campaigns. More recently, they have played a key role in helping to trace coronavirus exposure and slow the spread of the virus.

OCR, or Optical Character Recognition, is a process of recognizing text inside images and converting it into an electronic form. These images could be of handwritten text, printed text like documents, receipts, name cards, etc., or even a natural scene photograph.

OCR has two parts to it. The first part is **text detection** where the textual part within the image is determined. This localization of text within the image is important for the second part of OCR, **text recognition**, where the text is extracted from the image. Using these techniques together is how you can extract text from any image.

Table of Content

S No	Description	Page No
1.	Introduction of Project	4
2.	Related Work	5-6
3.	Methodology	7-10
4.	Proposed solution architecture	11-13
5.	Experimentation and Results	14-15
6.	Screenshots of working model	16-19
7.	Conclusion and Future Scope	20
8.	Project Links and References	21

Introduction

There are a number of desirable reasons to migrate from paper-based document system to paper-less document systems. A major reason being that, electronic records are easier to transmit and share remotely. They can also be easily stored and found. They reduce the cost of document storing and transferring, and can easily be backed up. Thus, the man-kind being the most intelligent species, that walked our planet, invented the concept of scanners. The first ever scanner was invented in 1957, by a team led by Russel A. Kirsch at the US National Bureau of Standards. It was developed for use with a computer, and was a drum scanner. An important concept, widely used in our paper is Computer Vision, which is a field in Artificial Intelligence, which deals with giving computers, high level understanding from digital images and videos. Computer vision deals with the automatic extraction, analysis and understanding of useful information from a single image or a series of images, also called as video. Digital image processing was pioneered at NASA's Jet Propulsion Lab in late 1960s to convert analogue signals to digital images with computer enhancement. For a computer, an image is a 2-D signal, made of rows and columns of pixels. In lame man's terms, in image processing, an image is processed when some transformation operations are performed on it. In technical terms, Image processing is usually related to the usage and application of mathematical functions and transformations over images. It is usually about an algorithm doing some transformations on the image such as stretching, contrasting, sharpening or smoothing on the image. The field of Computer Vision has high potential for further development in the future by implementing machine learning algorithms to better analyse the images.

The aim of this paper is to make use of Computer Vision and Python to implement the scanning application instead of introducing heavy and expensive hardware, we create a software that does the same with a simple camera, which will easily convert hard-paper documents into scanned images, to rotate, crop (if needed) and to save these documents in the directory of the user.

Apart from document scanning user gets the option of scanning a QR code or use text detection using OCR to convert image to text. Quick Response codes or in brief QR Codes, are two-dimensional (2D) matrix barcodes that are scanned using exploitation sensible and web capable smart mobile phones having camera, with QR Reader put in as default application, allows one to access some pre-written content such as a web site address, email address, details of things within the catalogue, phone numbers etc. QR Code is a reasonably 2-D symbology developed by Toyota subsidiary Denso Wave in 1994 with the first aim of being a symbol that is easily decoded by scanner instrumentation at high speed with additional knowledge content than conventional barcodes. Conventional Universal Product Code contains decoded data in one direction i.e. vertically into bars and house in between; whereas QR Code contains decoded data in each the directions i.e. vertical and horizontal direction. QR code is capable of holding additional volume of information than barcode, which is even a whole bunch of times as abundant data.

OCR or Optical Character Recognition has applications in a broad range of industries and functions. So, everything from scanning documents – bank statements, receipts, handwritten documents, coupons, etc. to reading street signs in autonomous vehicles falls under the OCR umbrella. Building an OCR system isn't a straightforward task. For starters, it is filled with problems like different fonts in images, poor contrast, multiple objects in an image, etc.

Related Work

QR code scanner

Data may be easily encoded in to QR codes by using any freely available QR code generators, enter the information to be encoded in to the sphere provided by generators. QR code generators may ask you to enter data to 1 to several Data Fields available, supported what data you're encoding. Once all the information fields crammed with necessary data in correct format QR code generator can publish the QR code for the data, which is able to be in the main in image format (JPEG, BMP, and PNG etc.). This could be used directly on internet or mails in e-format or may be revealed in print format. It's additionally attainable to disarrange the colour and even attainable to feature image in to QR codes to supply creative embellishment. To Cite: Shettar, I. M. (2016). Quick Response (QR) Codes in Libraries: Case study on the use of QR codes in the Central Library, NITK. In Proc. TIFR-BOSLA National Conference on Future Librarianship. (pp. 129-134).

Decoding of those QR codes may be done by on-line QR code decoders associated with any internet enabled good phone that has QR code Reader software package program pre-installed, if not the software may be downloaded from varied sites that are freely available on internet (Some of QR Code readers offered on-line are listed in later part). QR codes may be decoded with on-line decoder like ZXing Decoder on-line, MiniQR, on-line Barcode Reader, Saint Patrick Wied QR Generator, QR Code Generator and Recovery. QR codes can also be decoded with good Phone's Camera by inform towards QR code and scanning with image capture. Then pre-installed QR code reader decodes the QR code and displays content as text or uniform resource locator format. QR Codes additionally prompt your mobile device navigate to an internet page actions like dial variety, send SMS, Save Reminders, save variety to Phone Book etc.

Optical Character Recognition

Image binarization, which converts a grayscale image into black and white, is one of the most widely used preprocessing techniques for OCR since the 1980s. Otsu is a popular binarization method, which uses the grayscale histogram to find a threshold value. Object Attribute Thresholding uses grayscale and run-length histograms for unconstrained document images with complex backgrounds. O'Gorman proposed a global thresholding method using local connectivity details. Similarly, a five-step document binarization method focuses on connectivity issues in characters. Sauvola and Pietikainen proposed an adaptive binarization method which classifies the image into subcomponents and assigns different threshold values to pixels based on the component types. The double thresholding method developed by Chen et al. uses edge information to generate a binary image. Some more recent binarization techniques incorporate deep learning. For example, DeepOtsu applies Convolutional Neural Networks (CNN) iteratively, while Vo et al. adopt a CNN-based hierarchical architecture. Unknown-box Approx. to Improve OCR Performance 3 Skeletonization is another popular preprocessing technique, which aims to reduce the dimensions of an object. In the context of characters, skeletonization reduces the stroke thickness to 1-D curves. Similar to binarization, there are several existing methods for Skeletonization, some of which are specifically designed for OCR. Other preprocessing approaches exist in addition to binarization and skeletonization. Bieniecki et al. proposed methods to correct geometrical deformations in

document images. An independent component analysis-based method has been developed for handheld images of inscriptions. Harraj and Raissouni combined different image enhancement techniques with Otsu binarization. Deep learning-based Super-Resolution (SR) is employed in and Sporici et al. presented a CNN-based pre-processing method where convolution kernels are generated using Reinforcement Learning (RL).

Jaided AI, a company specializing in Optical Character Recognition Services, produces and maintains the EasyOCR package. It is able to write OCR texts in 70+ languages including English, Hindi, Russian, Chinese, and more.

Methodology

Scanning a document:

The steps that we need to follow to build this project are:

1. Convert the image to grayscale

OpenCV image processing `cv2.cvtColor()` method is used to change the colour of the image. More than 150 colour methods provided to change colour space. In this, we will change RGB to Gray colours space

2. Reducing noise and detecting the edges in the image

The most important part for detecting an image accurately is to remove the unnecessary noise present in it. there are many methods provided in OpenCV. Clearing noise of image with (`cv2.erode`, `cv2.morphologyEx`) and edges can be found by `cv2.Canny` method.

3. Image thresholding and finding contours

Thresholding helps us in archiving better result in finding Contours of image. Contours are nothing but the edges region this could be achieved by `cv2.adaptiveThreshold` and `cv2.findContours` method in OpenCV.

4. Find the biggest contour area and approx poly

Next step is to find the biggest contour and crop the image to that contour and display it. for counting the biggest contour out of all counters OpenCV provides `cv2.contourArea()` method and after getting the main contour we have to find the approx-poly main coordinates the image by (`cv2.approxPolyDP`, `cv2.arcLength`).

5. Apply warp perspective to get the top-down view of the document

The final step is to crop the image to approx-poly with (`cv2.getPerspectiveTransform`, `cv2.warpPerspective`) and display it.

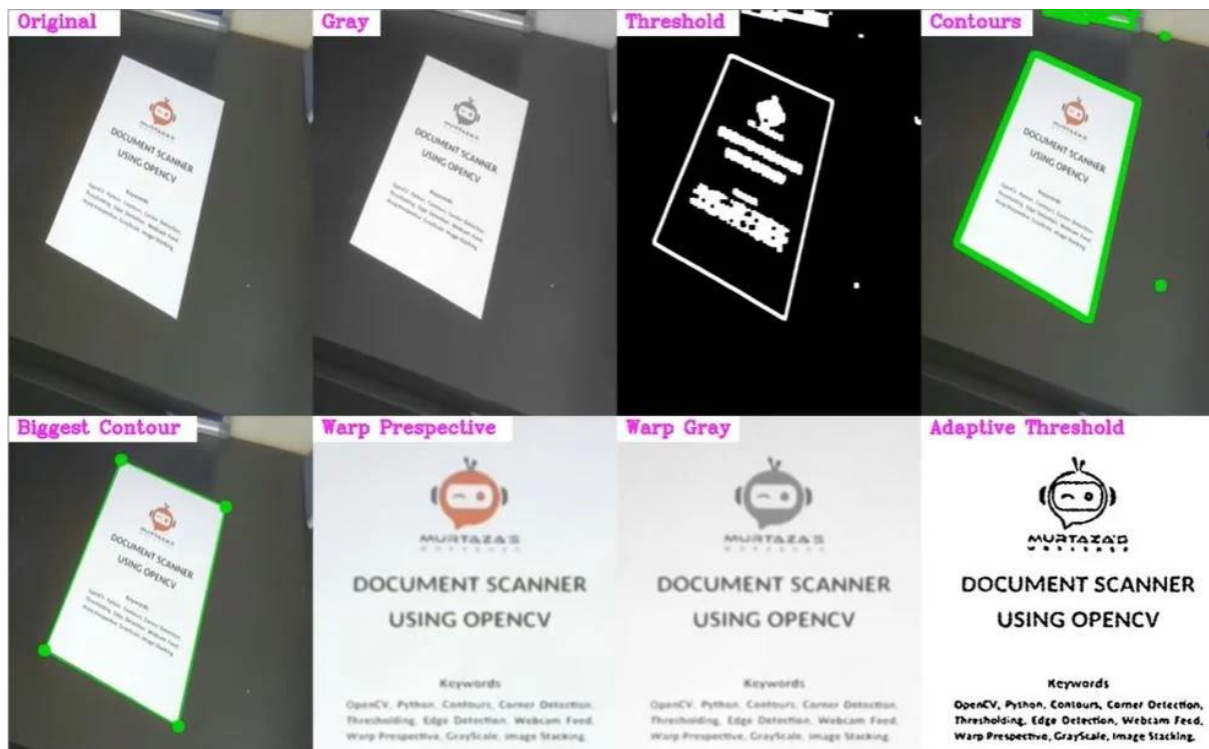


Fig: Process of Scanning the document

Scanning a QR code:

A standard QR code has the following components:

1. **Quiet Zone** - This is the empty white border around the outside of a QR code. Without this border, a QR reader will not be able to determine what is and is not contained within the QR code (due to interference from outside elements).
2. **Finder pattern** - QR codes usually contain three black squares in the bottom left, top left, and top right corners. These squares tell a QR reader that it is looking at a QR code and where the outside boundaries of the code lie.
3. **Alignment pattern** - This is another smaller square contained somewhere near the bottom right corner. It ensures that the QR code can be read, even if it is skewed or at an angle.
4. **Timing pattern** - This is an L-shaped line that runs between the three squares in the finder pattern. The timing pattern helps the reader identify individual squares within the whole code and makes it possible for a damaged QR code to be read.
5. **Version information** - This is a small field of information contained near the top-right finder pattern cell. This identifies which version of the QR code is being read (see "Types of QR code" below).
6. **Data cells** - The rest of the QR code communicates the actual information, i.e., the URL, phone number, or message it contains.
7. **Error correction level**- It tells about the percentage of data bytes that can be recovered. Based on the level of error correction, there can be 4 different categories. Low(7%), medium(15%), quartile(25%) and high(30%). The higher the level, the lower is the storage capacity.

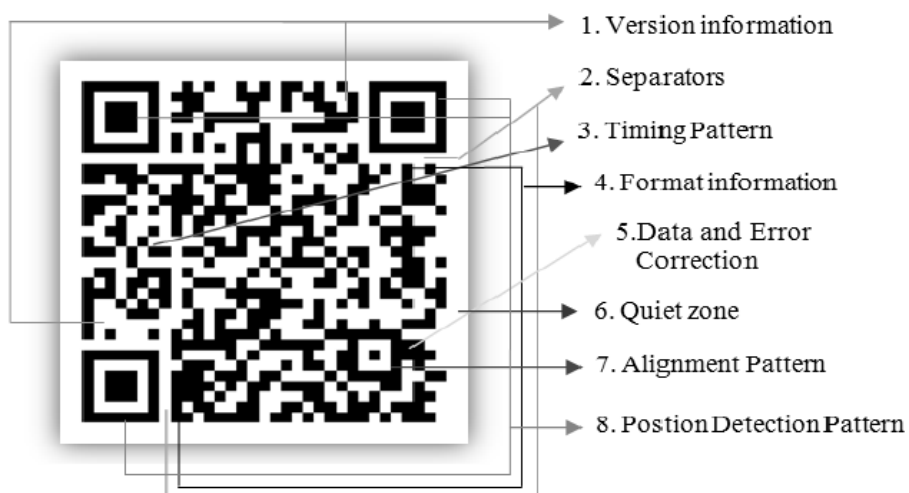


Fig: Structure of a QR code

Text recognition:

OCR, or Optical Character Recognition, is a process of recognizing text inside images and converting it into an electronic form. These images could be of handwritten text, printed text like documents, receipts, name cards, etc., or even a natural scene photograph.

OCR has two parts to it. The first part is **text detection** where the textual part within the image is determined. This localization of text within the image is important for the second part of OCR, **text recognition**, where the text is extracted from the image. Using these techniques together is how you can extract text from any image.

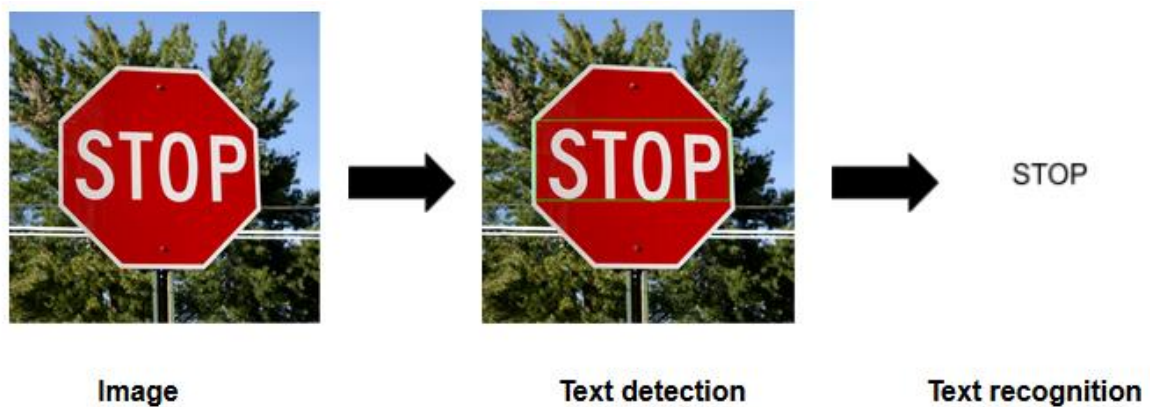


Fig: Process of Optical Character Recognition

Model Architecture

Scanning a document:

In the proposed system, we need to process our image, for which we need to take the input of the image first. In this case, the image can be directly read into the scanner application using OpenCV. The scanned image is then converted to grayscale. Contours of this selected image are found. (A contour is a closed loop joining all the continuous points, having same intensities). A python list is created to store the Area(s) of all the contour(s) found and a python dictionary is created to store the calculated area(s) of the contour(s) with their corresponding contour index value(s). Then to find the object of interest, which is to be scanned, in our input image. We find the contour of maximum area and its index. Next, we find a rectangle of minimum area that fits the contour selected, that is, maximum area contour in the image. We use OpenCV's function to find the co-ordinates of this rectangle so formed. The points are in the order, [top left, top right, bottom left, bottom right]. The points thus found, are passed to find the perspective transform of the image, along with another set of points, which decide the alignment/positioning of the object of interest in the final image. We also create a brightened copy of this transformed image by adding '50' to each pixel of the image obtained. The image thus found after this is sharpened and any unwanted noise in the image is removed, using various filters, like Gaussian filter, which is used to remove noise in an image. A copy of this image is converted to grayscale. Threshold of the grayscale image is calculated and bitwise not operator is applied on this image, to change the black to white and white to black again and is sharpened. Adaptive threshold is applied to another copy of the gray perspective transformed image. Bitwise and is performed between these two images to get the resulting image. The resultant image is displayed, under the title of "Scanned image".

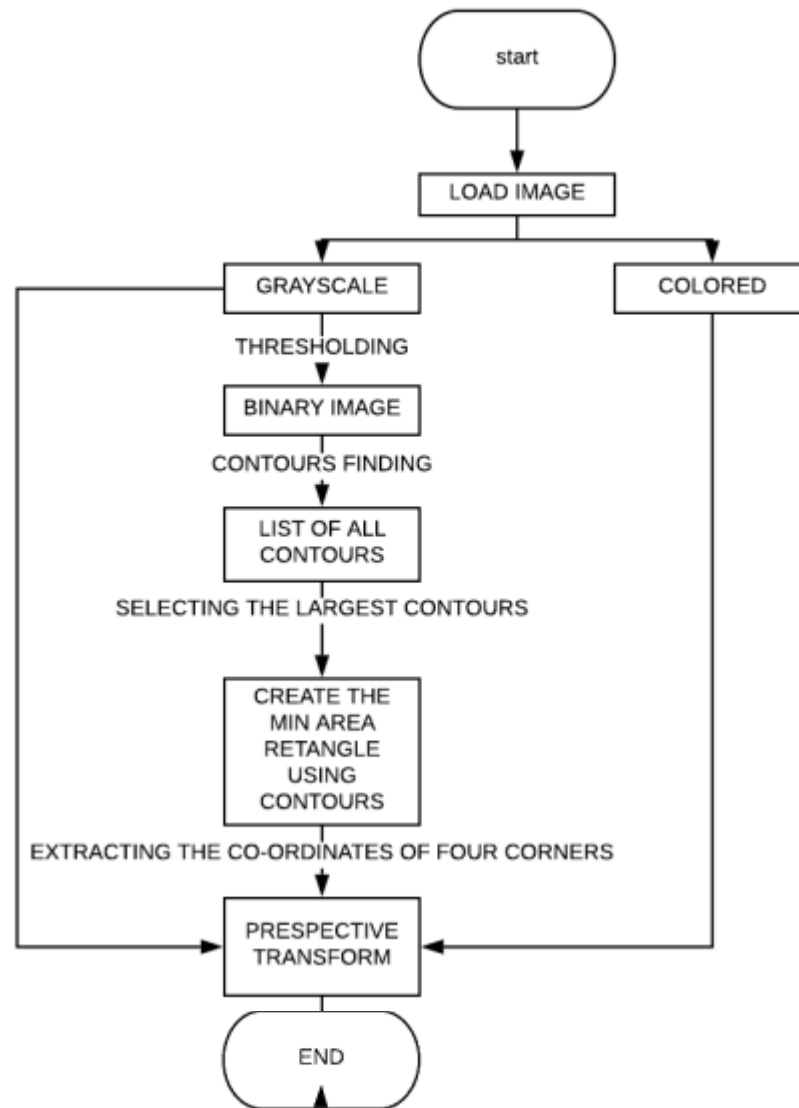


Fig: Architecture of Scanning process

Scanning a QR code:

We design a QR code recognition procedure for decoding QR code accurately and rapidly. First, we propose a new local binarization method to overcome images with uneven light and nonuniform background problems. Binarization is an important process for accurately recognizing black-and-white module in QR code images. QR code extraction is responsible for searching QR code area in images. QR code extraction consists of searching finder patterns, estimating version, and searching alignment pattern. We use the proportion of finder patterns to locate the QR code position. After searching three finder patterns, we use the distance between these patterns to estimate version. Finally, we develop a connected-component method to search alignment pattern accurately. After locating QR code position, we resample QR code image by perspective transformation. Finally, we get the corrected data by error correction algorithm.

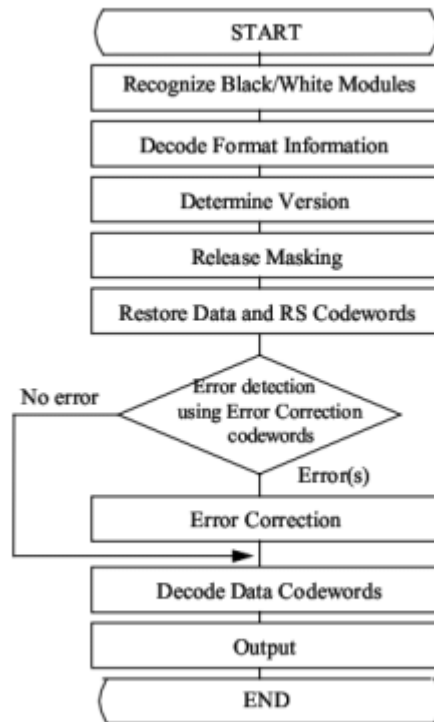


Fig: Architecture of QR scanning process

Text detection:

EasyOCR is built with Python and Pytorch deep learning library, having a GPU could speed up the whole process of detection. The detection part is using the CRAFT algorithm and the Recognition model is CRNN. It is composed of 3 main components, feature extraction (we are currently using Resnet), sequence labelling (LSTM) and decoding (CTC). EasyOCR doesn't have many software dependencies, it can directly be used with its API.

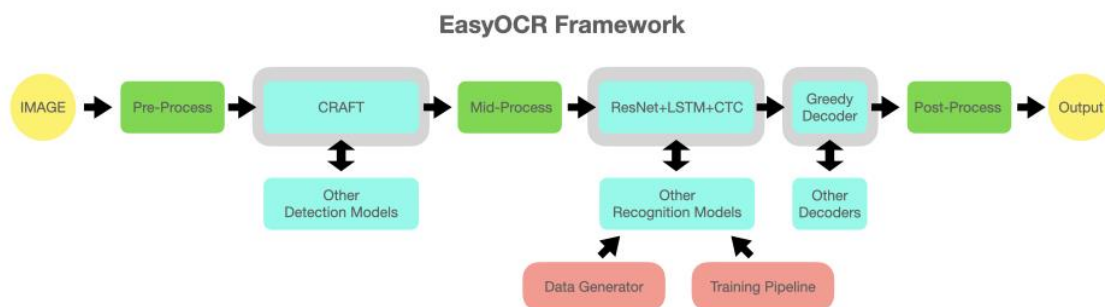
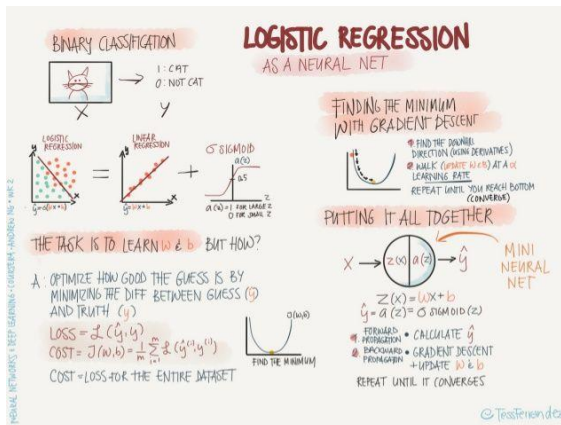


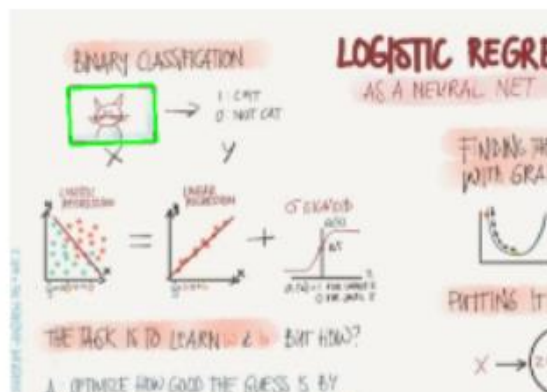
Fig: Architecture of Text detection process

Experimentation and Results

1. Unbounded documents: Documents in which edges can't be detected properly give wrong results.



Input Document



Scanned Output



Download
Upload an image

2. Problem in scanning QR codes: The quiet zone that distinguishes the QR Code from the surroundings is too small or nonexistent, so the QR Code can't be read. Because some designs have left out the quiet zone, the scanner can't determine what is the graphic and what is the QR Code.

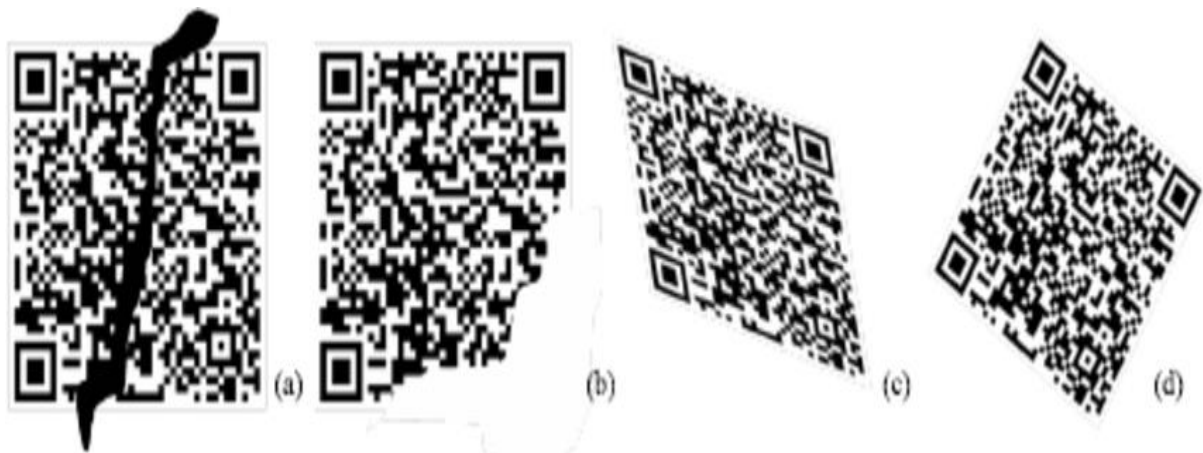


Fig: Samples of QR Code dirty (a), damaged (b), distorted (c), and rotated(d).

3. Comparing two popular OCR methods and choosing the better one according to the task: EasyOCR is lightweight and does a better job on numbers. If your document is alphabet-heavy, you may give Tesseract higher weights. A hybrid process may be considered for more accurate results. When it comes to speed, Tesseract is more favourable on a CPU machine, but EasyOCR runs extremely fast on a GPU machine.

Screenshots of my project

1. Scanning a document

Document/QR code Scanner

Would you like to scan a document or QR code?

Document

Upload the image to scan your: Document

Upload your picture



Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG

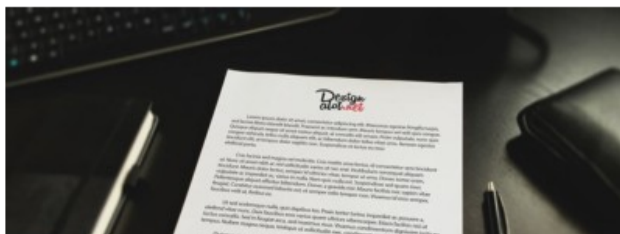
Browse files



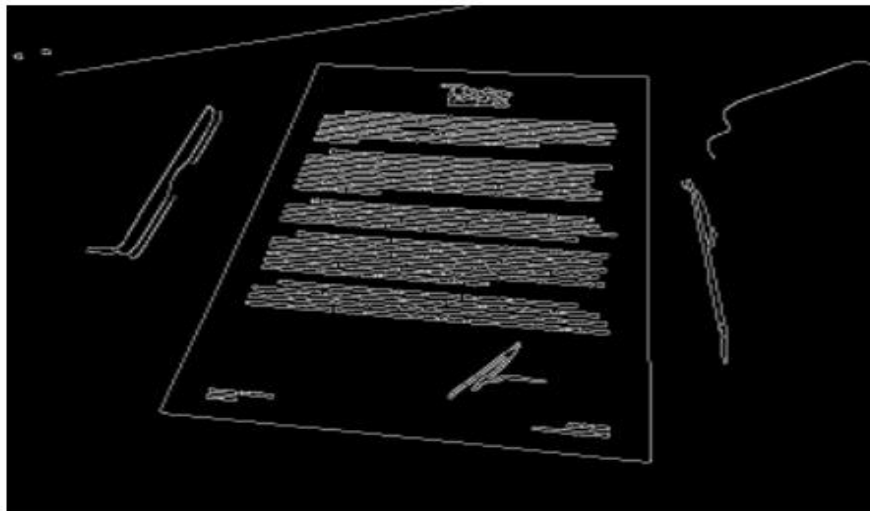
document_curve.jpg 86.4KB



Successfully Uploaded!



Output



Finding contour...



Almost Ready...



Download

Upload an Image

2. Text Detection using OCR

Document/QR code Scanner

Would you like to scan a document or QR code?

Text Detection

Upload the image to scan your: Text Detection

Upload your picture



Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG

Browse files



download.png 1.5MB



Successfully Uploaded!



```
[
  0 : "Keep"
  1 : "Learning"
  2 : "Keep"
  3 : "Growing"
]
```

3. Decoding a QR Code

Document/QR code Scanner

Would you like to scan a document or QR code?

QR code

Upload the image to scan your: QR code

Upload your picture



Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG

Browse files

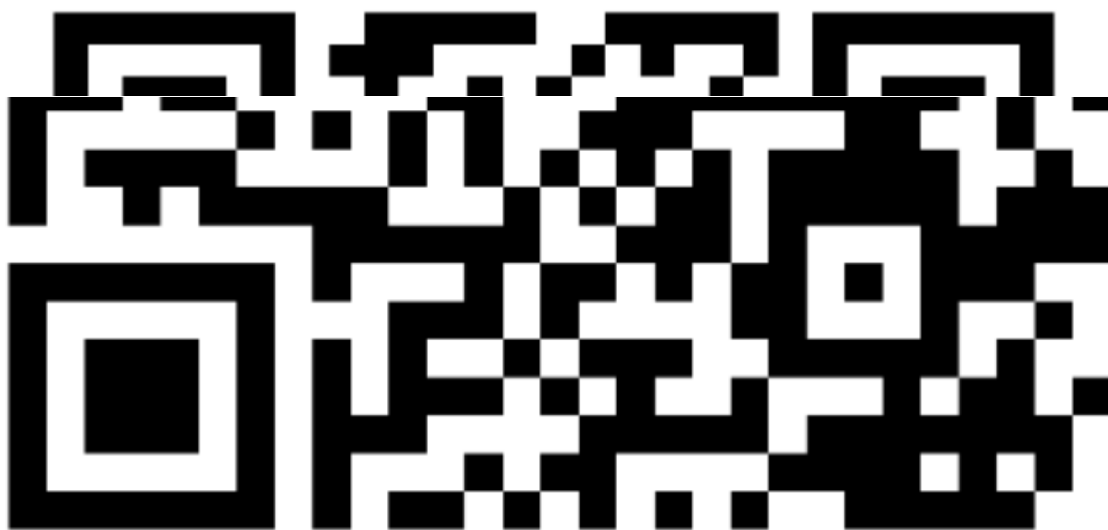


sargun qr.png 44.3KB



Successfully Uploaded!

Original Image



Original Image

Decoded data

Decoded data: <https://www.linkedin.com/in/sargun-bindra-19209a194/>

Upload an Image

Future Scope and Conclusion

Many industries are using QR Codes already. These include textiles, real estate, manufacturing, IT, etc. And many of them are looking forward to leveraging the QR Code technology to level up the scale of their operations and promotions. In the era of social distancing, contactless operations have become the norm. From contactless payments to promotions, QR Coded will be seen everywhere.

OCR is used for handwriting recognition tasks to extract information. A lot of work is going on in this field and we have made some really significant advancements. Microsoft has come up with an awesome mathematical application that takes as input a handwritten mathematical equation and generates the solution along with a step-by-step explanation of the working.

OCR is increasingly being used for digitization by various industries to cut down manual workload. This makes it very easy and efficient to extract and store information from business documents, receipts, invoices, passports, etc. Also, when you upload your documents for KYC (Know Your Customer), OCR is used to extract information from these documents and store them for future reference.

OCR is also used for book scanning where it turns raw images into a digital text format. Many large scale projects like the Gutenberg project, Million Book Project, and Google Books use OCR to scan and digitize books and store the works as an archive. The banking industry is also increasingly using OCR to archive client-related paperwork, like onboarding material, to easily create a client repository. This significantly reduces the onboarding time and thereby improves the user experience. Also, banks use OCR to extract information like account number, amount, cheque number from cheques for faster processing.

The applications of OCR are incomplete without mentioning their use in self-driving cars. Autonomous cars rely extensively on OCR to read signposts and traffic signs. An effective understanding of these signs makes autonomous cars safe for pedestrians and other vehicles that ply on the roads.

There are definitely many more applications of OCR like vehicle number plate recognition, converting scanned documents into editable word documents, and many more.

Github Link and App Link

Code Link :- <https://github.com/sargun06/DocumentScanner>

Link of the Webapp:- <https://share.streamlit.io/sargun06/documentscanner/main/main.py>

References

1. <https://github.com/JaidedAI/EasyOCR>- The github link of easyocr package which is being used for text detection.
2. Boris Rewald, of University of Natural Resources and Life Sciences, did a comparative study of “Scanner baseband Camera based MR system or Mini Rhizotron Imaging system”, February (2019).
3. Aleksey D. Poudalov, Michael IV Piltsov, Vladimir G. Mazur Angarsk of state technical university, Russian Fed., “Algorithmization of the Spatial Method of Image Enhancement Based on Aligning the BrightnessHistograms”, 2018 14th International Scientific-Technical Conference. (2018)
4. Li Xin and Shi YiLiang of Wuhan University of Science and Technology, “Computer Vision Imaging based on Artificial Intelligence”, 2018 International Conference on Virtual Reality and Intelligent Systems. (2015)
5. M. Sridevi, National Institute of Technology, Tiruchirappalli, India, “Image Segmentation based on Multilevel Thresholding using Firefly Algorithm”, Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017), IEEE Xplore Compliant. (2016)
6. Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multiperson 2d pose estimation using part affinity fields. In CVPR, pages 1302–1310. IEEE, 2017.
7. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. PAMI, 40(4):834–848, 2018.